A prototypical cytidylyltransferase: CTP:glycerol-3-phosphate cytidylyltransferase from *Bacillus subtilis*

Christian H Weber[1,2†], Young Seo Park[2‡], Subramaniam Sanker[2], Claudia Kent[2] and Martha L Ludwig[1,2*]

Background: The formation of critical intermediates in the biosynthesis of lipids and complex carbohydrates is carried out by cytidylyltransferases, which utilize CTP to form activated CDP-alcohols or CMP-acid sugars plus inorganic pyrophosphate. Several cytidylyltransferases are related and constitute a conserved family of enzymes. The eukaryotic members of the family are complex enzymes with multiple regulatory regions or repeated catalytic domains, whereas the bacterial enzyme, CTP:glycerol-3-phosphate cytidylyltransferase (GCT), contains only the catalytic domain. Thus, GCT provides an excellent model for the study of catalysis by the eukaryotic cytidylyltransferases.

Results: The crystal structure of GCT from *Bacillus subtilis* has been determined by multiwavelength anomalous diffraction using a mercury derivative and refined to 2.0 Å resolution (R_{factor} 0.196; R_{free} 0.255). GCT is a homodimer; each monomer comprises an α/β fold with a central 3-2-1-4-5 parallel β sheet. Additional helices and loops extending from the α/β core form a bowl that binds substrates. CTP, bound at each active site of the homodimer, interacts with the conserved ^{14}HXGH and ^{113}RTXGISTT motifs. The dimer interface incorporates part of a third motif, ^{63}RYVDEVI, and includes hydrophobic residues adjoining the HXGH sequence.

Conclusions: Structure superpositions relate GCT to the catalytic domains from class I aminoacyl-tRNA synthetases, and thus expand the tRNA synthetase family of folds to include the catalytic domains of the family of cytidylyltransferases. GCT and aminoacyl-tRNA synthetases catalyze analogous reactions, bind nucleotides in similar U-shaped conformations, and depend on histidines from analogous HXGH motifs for activity. The structural and other similarities support proposals that GCT, like the synthetases, catalyzes nucleotidyl transfer by stabilizing a pentavalent transition state at the α-phosphate of CTP.

Addresses: [1]Biophysics Research Division, University of Michigan, Ann Arbor, MI 48109, USA and [2]Department of Biological Chemistry, University of Michigan, Ann Arbor, MI 48109, USA.

Present addresses: [†]Pathology Department, University of Michigan Medical School, Ann Arbor, MI 48109, USA and [‡]Department of Food and Bioengineering, College of Bioengineering, Bokjung-Dung, Sujung-Ku, Sungnam, South Korea.

*Corresponding author.
E-mail: ludwig@biop.umich.edu

Key words: CTP binding, cytidylyl transfer, substrate activation, X-ray structure

Received: 18 February 1999
Revisions requested: 29 March 1999
Revisions received: 30 April 1999
Accepted: 27 May 1999

Published: 27 August 1999

Structure September 1999, 7:1113–1124
http://biomednet.com/elecref/0969212600701113

0969-2126/99/$ – see front matter

Introduction

Cytidylyltransferases activate intermediates for biosynthesis in a variety of biochemical reactions. In particular, cytidine triphosphates are utilized specifically to form key intermediates in the major pathways of lipid biosynthesis. The best-characterized eukaryotic cytidylyltransferase is CTP:choline-phosphate cytidylyltransferase (CCT), which catalyzes the activation of phosphocholine (Equation 1) for the synthesis of phosphatidylcholine.

$$\text{CTP} + \text{choline-phosphate} \rightleftharpoons \text{CDP-choline} + \text{pyrophosphate} \quad (1)$$

In addition to its structural role in forming the membrane bilayer, phosphatidylcholine is a component of serum lipoproteins and is a substrate for enzymes of signal transduction pathways [1]. CCT regulates the overall biosynthesis of phosphatidylcholine, and the activity of CCT is modulated in response to the needs of the cell for phosphatidylcholine. Mammalian CCT is highly phosphorylated and is activated by lipids [2]; levels of activity correlate with phosphorylation and membrane association [2–4]. Elucidation of the function and mechanism of CCT is important for understanding not only phosphatidylcholine biosynthesis *per se* but also the roles of this enzyme in vesicular trafficking [5] and apoptosis [6]. However, the heterogeneity of phosphorylation [7] and the association with lipids and membranes make intact CCT and other eukaryotic cytidylyltransferases difficult subjects for structure analysis.

To understand the mechanisms and regulation of catalysis in cytidylyltransferases such as CCT, we have begun structure-function studies with the closely related bacterial cytidylyltransferase CTP:glycerol-3-phosphate cytidylyltransferase (GCT) from *Bacillus subtilis*. As shown in Equation 2, GCT activates glycerol-3-phosphate; the product CDP-glycerol is utilized in the biosynthesis of

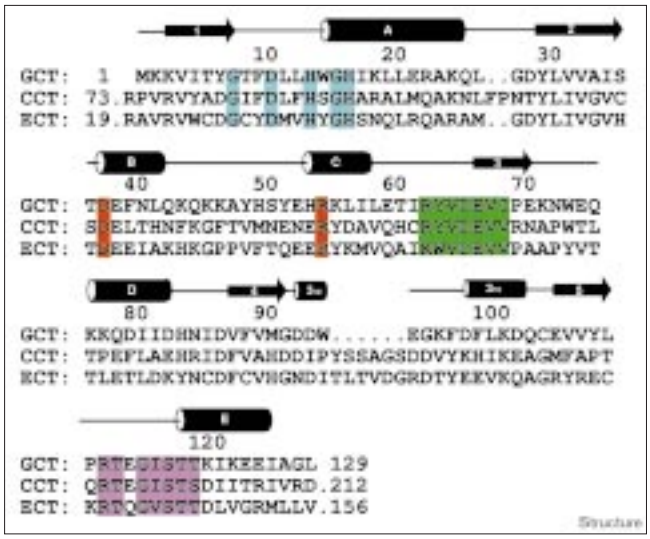
teichoic acid, a component of the cell wall of *B. subtilis* and some other gram-positive bacteria.

$$\text{CTP} + \text{glycerol-3-phosphate} \rightleftharpoons \text{CDP-glycerol} + \text{pyrophosphate} \quad (2)$$

The amino acid sequence of GCT is very similar to that of CCT and the eukaryotic CTP:ethanolamine phosphate cytidylyltransferase (ECT), as shown in Figure 1. CCT, ECT and GCT thus constitute a family of proteins [8] that we call the GCT family. GCT is the smallest of these cytidylyltransferases: the catalytic domain comprises the entire sequence. ECT functions in phosphatidylethanolamine biosynthesis, and actually contains two copies of the catalytic domain per polypeptide chain, although the second copy is less conserved than the first [9,10]. The sequence alignments in Figure 1 show that ~30% of the residues in GCT from *B. subtilis* are identical to residues in the catalytic domains of the larger members of the GCT family. This level of sequence identity implies that GCT will provide a good structural model for the catalytic domains from the larger eukaryotic cytidylyltransferases CCT and ECT.

Several regions of the aligned sequences of the GCT family members are particularly striking [8]. The first is the sequence ^{8}GX(Y/F)DXXHXGH (in single-letter amino acid code). The ^{14}HXGH portion of this sequence is a conserved motif in another nucleotidyltransferase family, the class I aminoacyl-tRNA synthetases [11]. In this class of enzymes the histidine residues of the HXGH sequence have been shown to stabilize the transition state [12], and appear to bind the α-phosphate of the ATP moiety of the transition state [13]. Mutation of the corresponding His14 and His17 residues of GCT decreases k_{cat} by factors of 10^4 and 10^3, respectively [8]. The second highly conserved sequence in the GCT family is ^{113}RTXGISTT. This motif has not been identified in other enzyme families, and thus is a signature that distinguishes the GCT family of cytidylyltransferases. A third motif that is conserved throughout the GCT family is ^{63}RYVDEVI.

Figure 1



Alignment of the sequence of GCT with the catalytic domains of two other related cytidylyltransferases, choline-phosphate cytidylyltransferase (CCT) from rat (residues 73–212), and the first domain of human ethanolamine-phosphate cytidylyltransferase (ECT; residues 19–156). The secondary structure elements observed in GCT are indicated above the sequences: arrows denote β strands and cylinders denote helices. Conserved sequence motifs are color-coded: the ^{8}GXXDXXHXGH sequence is in cyan, the ^{113}RTEGISTT loop is in magenta, and the ^{63}RYVDEVI sequence is in green. Asp38 and Arg55 are shown in red. The color scheme for the conserved sequences is preserved throughout the figures.

In this paper we report the determination of the crystal structure of GCT by multiwavelength anomalous diffraction (MAD), and describe the features of the molecular model that has been obtained by refinement. The structure provides essential information about the roles of the signature sequences and other conserved residues. In particular, the crystalline enzyme contains tightly bound CTP at the active sites, and the ^{14}HXGH and ^{113}RTXGISTT sequences are closely associated with the bound substrate.

Results and discussion

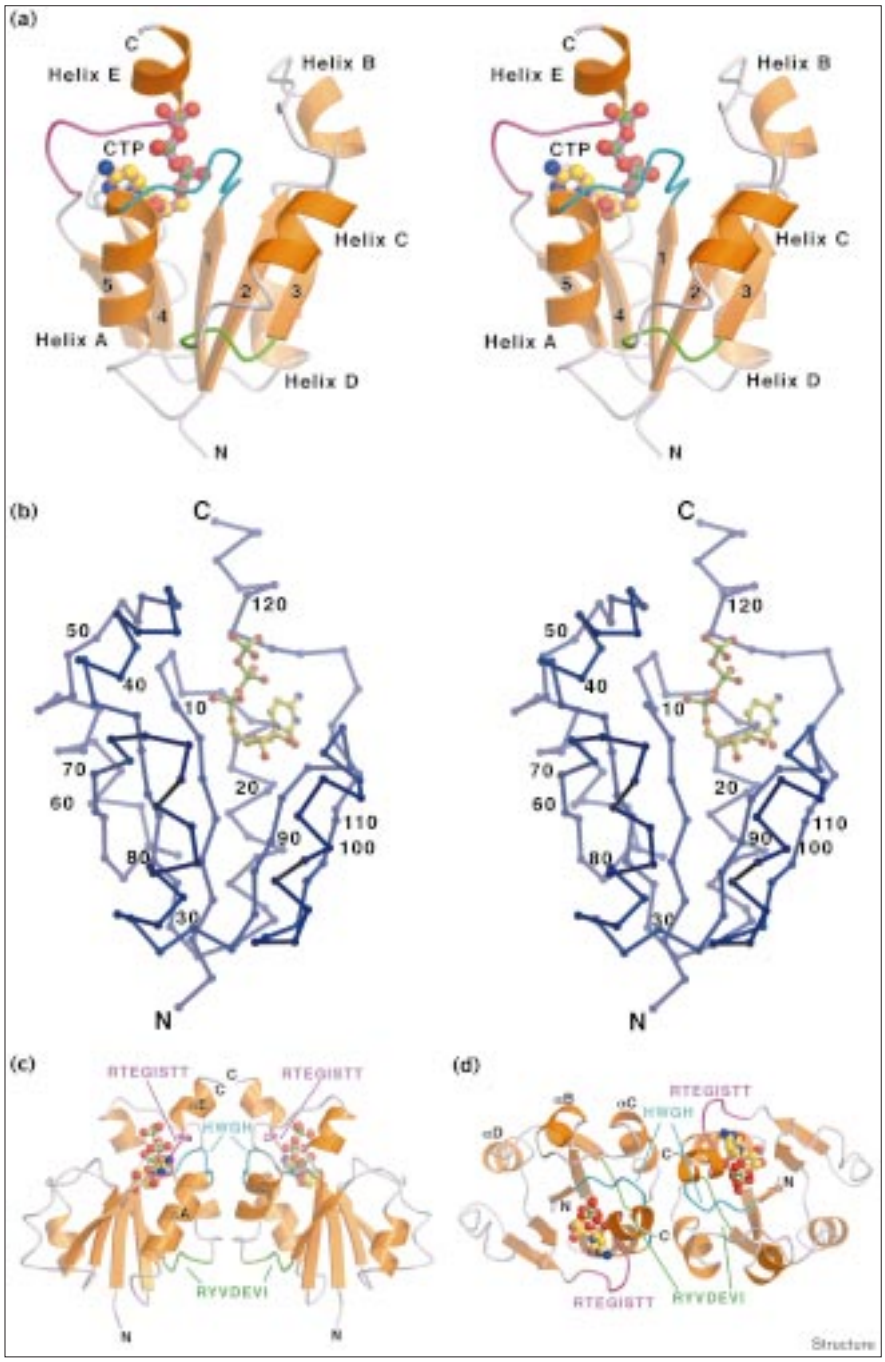
Structure determination

GCT was crystallized in two space groups: an R3 rhombohedral form that diffracts to 2.7 Å and a $P2_1$ monoclinic form that diffracts to 2.0 Å, but displays variable degrees of merohedral twinning. In both crystal forms a dimer of GCT constitutes the asymmetric unit. Experimental phases were determined for the R3 form by MAD using a mercuric iodide derivative with Hg bound at Cys106. The monoclinic structure was subsequently solved by molecular replacement and was refined using X-PLOR (as described in the Materials and methods section). The current model (R_{factor} = 0.196, R_{free} = 0.255) includes 252 of the 258 residues in the dimer, two bound CTP molecules and 114 solvent molecules.

Description of the monomer fold and the dimer interface

The structure of the monomer of GCT is illustrated in Figures 2a,b. Five parallel β-sheet strands with topology 3-2-1-4-5 are flanked by helices to form a doubly wound nucleotide-binding domain [14]. Three α helices, A, C and D, pack against the sheet in accord with the right-hand rule for α/β supersecondary structures. However, the cross-connector between strands 4 and 5 (residues 94–104), at the periphery of the monomer and exposed to solvent, does not form a regular α helix, but instead consists of short 3_{10} helices interrupted by β turns. We will refer to this part of the structure as 'the 3_{10} segment'. Helices B and E extend above the β sheet and, along with the loops carrying the conserved motifs ^{8}GTFDLLHWGH and ^{113}RTEGISTT, form a bowl that binds the substrate CTP

Figure 2

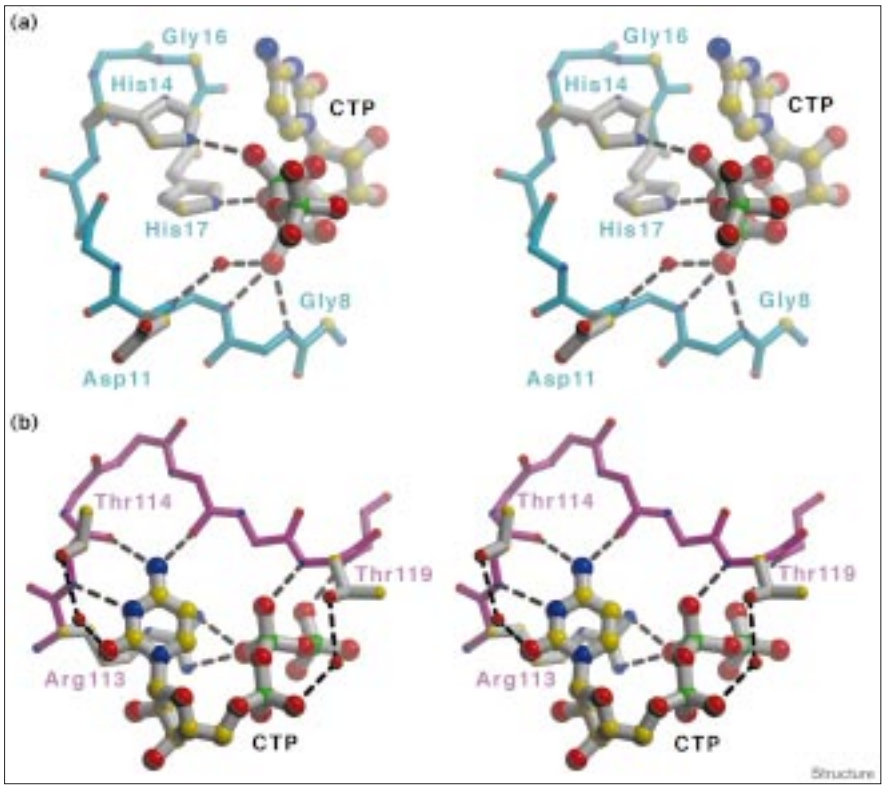


The structure of GCT. (a) Stereoview ribbon representation of the GCT monomer with β strands and α helices labeled. CTP, bound at the C terminus of the β sheet, is shown in ball-and-stick representation. The ^{14}HWGH motif (cyan) is located at the start of helix A and the ^{113}RTEGISTT loop (magenta) follows strand 5 of the sheet. Despite its distance from the active site, the region ^{63}RYVDEVI (green) is also highly conserved. (b) Stereoview Cα trace of the GCT monomer rotated about 180° from the view in (a). (c,d) Ribbon diagrams of the GCT dimer showing the structural elements that contribute to the dimer interface and the locations of the three conserved motifs. The two views are related by a 90° rotation about a horizontal axis. In (c) the twofold noncrystallographic symmetry axis is vertical; in (d) the view is along the twofold axis.

(see below). The third conserved sequence, ^{63}RYVDEVI (Figure 1), is found at the 'bottom' of the monomer distant from the CTP-binding site.

Side and top views of the dimer are shown in Figures 2c,d. The pair of chains related by the noncrystallographic twofold axis is assumed to correspond to the homodimer observed in solution by gel filtration [15]; no other pairing around a simple dyad is possible in the observed crystal forms. The area and properties of the interface support this assignment. An accessible area of 1050 Å^2 from each chain, or 14.8% of the total accessible surface, is buried at the dimer interface and the fraction of nonpolar residues is 65%, close to the average for dimer interfaces [16]. One of the residues buried by dimer formation is Arg63 from the ^{63}RYVDEVI sequence; the

Figure 3



Detailed views of the residues implicated in substrate-binding and catalysis. (a) Stereoview of the interactions between the ^{14}HWGH motif and CTP. His14 and His17 hydrogen bond (dashed lines) to the β- and α-phosphate oxygens, respectively. (b) Stereoview of the interactions between the ^{113}RTEGISTT motif and CTP, showing the U-shape of the bound CTP. Nucleotide specificity is conferred through backbone interactions of Thr114 and Ile117 (see text).

hydrophobic segments of the arginine sidechains from each monomer interact around the local twofold axis. In mammalian CCT, an Arg→His mutation at the site corresponding to Arg63 results in a temperature-sensitive enzyme [17]. Other residues of the interface include Trp15 of the ^{14}HWGH motif, and the two hydrophobic residues (Leu12 and Leu13) that precede ^{14}HWGH. The substrate-binding sites are separated by a wall formed by helices C and E and by the twofold related ^{14}HWGH motifs. Residues that interact with CTP do not extend across the dimer interface, suggesting that each chain possesses a self-contained active site. However, because residues adjoining the ^{14}HWGH sequence that binds CTP are part of the interface, interactions between the two active sites of the dimer are possible.

The 3-2-1-4-5 topology of the β sheet occurs in many other α/β proteins, including the class I aminoacyl-tRNA synthetases. Prior to the determination of the crystal structure, the structure of GCT was modeled by Bork and colleagues using a threading algorithm [18]. Their prediction, based primarily upon the sequence homology between GCT and the class I aminoacyl-tRNA synthetases, correctly placed the central five-stranded β sheet, four of the five α helices, and the HXGH motif. More detailed comparisons between GCT and the aminoacyl-tRNA synthetases are presented below.

CTP binding

A cleft between the C termini of β strands 1 and 4, at the cross-over of the parallel β sheet [19], serves as the base of a bowl that binds the substrate CTP (Figure 2). In each monomer the bound CTP adopts a horseshoe or U-shape with the ribose protruding most deeply into the β sheet. The cytidine base is *anti* to the ribose, which adopts the *2′ endo* conformation. The horseshoe conformation orients the substrate for in-line nucleophilic attack at the α-phosphate. In contrast, the kinases and phosphoryl transferases, which catalyze transfer of the γ-phosphate, often bind nucleoside triphosphates in more extended conformations — adenylate kinase [20,21] and ATP synthase [22] are prime examples. The U-shape is not a requirement for nucleotidyl transfer, however; in DNA polymerases [23] and in kanamycin transferase [24] the bound nucleotides adopt other conformations that expose the α-phosphate to in-line attack.

Residues of the ^{14}HWGH motif and the ^{113}RTEGISTT loop dominate the interactions that are responsible for binding CTP; these interactions are depicted in Figures 3a,b and shown schematically in Figure 4. ^{15}WGH forms the initial turn of helix A, which is capped by His14. The imidazole sidechain of this histidine hydrogen bonds to NH16 and to the β-phosphate oxygen of CTP; His17 and the backbone NH of Phe10 interact with the α-phosphate

oxygens. An unusual backbone conformation, where the amides of residues Thr9, Phe10 and Asp11 all point towards an α-phosphate oxygen, can be seen in Figure 3a. The conserved Gly8 residue allows close approach of the mainchain to the ribose of CTP.

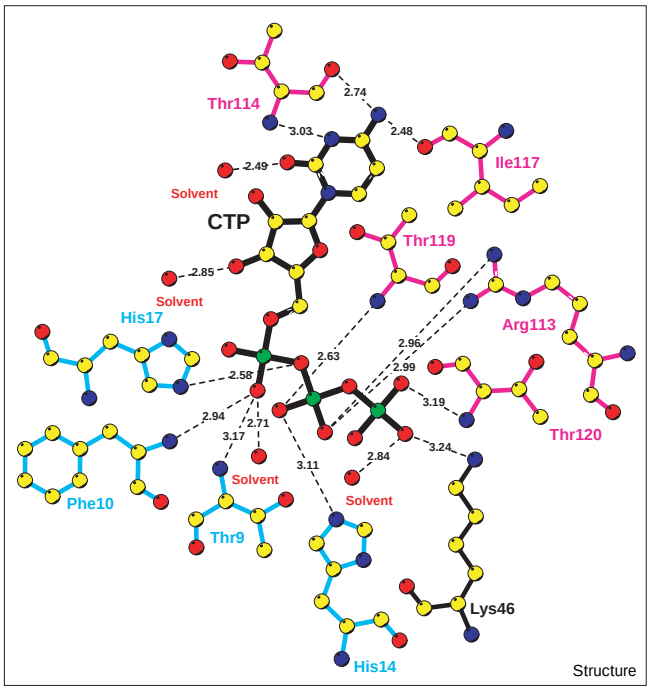
^{113}RTEGISTT is a characteristic fingerprint sequence for the GCT family of cytidylyltransferases. In the GCT–CTP complex Glu115 and Gly116 are the central residues of a type II turn; Thr119 and Thr120 are the initial residues of helix E. The negative charges of the CTP triphosphate are partly stabilized by the dipole of helix E and by the positively charged Arg113. Arg113 stacks against the nucleotide ring and interacts with a β-phosphate oxygen (Figure 3b). Oxygen atoms of the β- and γ-phosphates of CTP hydrogen bond to the backbone amides of Thr119 and Thr120 from helix E. Helix E is not part of the central α/β sandwich (Figure 2a) and the interaction of the nucleotide phosphates with helix E is a special feature of GCT and glutaminyl-tRNA synthetase, which distinguishes these enzymes from the kinases. Furthermore, the binding site for the α- and β-phosphates of CTP, near the start of helix A, is not spatially equivalent to the sites formed by the P loops of kinases [25–27] or by the GXXGXG fingerprints found in the nucleotide-binding folds of enzymes such as lactate dehydrogenase [28–30].

Nucleotide specificity is conferred chiefly through backbone interactions formed between the turn residues Thr114 and Ile117 and the pyrimidine base. Carbonyl oxygens of both these residues form hydrogen bonds with the exocyclic 4-NH_2 group, and the backbone NH of Thr114 donates a hydrogen bond to the N3 nitrogen. This pattern of hydrogen bonding presumably is sufficient to distinguish CTP from UTP, in which the 4-NH_2 group is replaced by an oxygen atom and N3 is protonated. The nucleotide-binding site is too small to permit binding of the purine triphosphates in the CTP pocket. In contrast to the nucleotide base, the ribose moiety has comparatively few contacts with the protein. The amide nitrogen of Gly92 and a solvent molecule interact with the 3′-oxygen. The 2′-oxygen does not form hydrogen bonds in the crystal structure, which is consistent with the finding that the activity of the enzyme with dCTP is 95% of that with CTP [15]. In fact, dCTP is also a good substrate for mammalian CCT [31].

Mg^{2+} is essential for cytidylyltransferase activity and presumably interacts with the phosphates of bound substrates. The crystals we have analyzed, however, were grown from protein purified in the presence of EDTA, and we cannot identify the sites that bind Mg^{2+}.

Bound CTP was observed in the crystal structure, even though the enzyme had been dialyzed in the absence of CTP before crystallization, suggesting a much higher affinity of the enzyme for CTP than would be indicated by the very large K_m for CTP (1.4 mM). Preliminary substrate-binding studies show that the K_D for CTP binding to the enzyme to form the binary enzyme–CTP complex is indeed much smaller than the K_m [32].

Figure 4

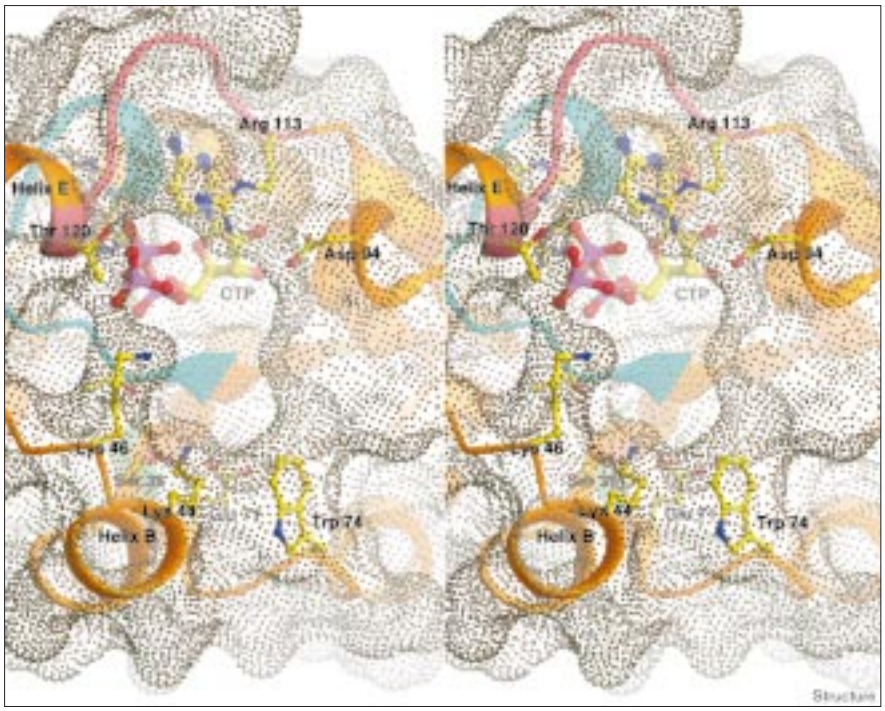


Schematic drawing of the CTP ligand and its hydrogen-bonding interactions with residues of GCT. Bond distances are given in Å.

The active-site bowl

The CTP molecule is bound in a large cavity, or bowl, that is best visualized in Figure 5. One side of the substrate bowl is formed from the conserved ^{8}GTFDLLHWGH and ^{113}RTEGISTT sequences; the other side includes residues from helix B, the loop following strand 3, and the 3_{10} segment. An empty pocket next to the CTP is the likely binding site for glycerol-3-phosphate, which can be docked in this region. Model building suggests that in-line nucleophilic attack on the α-phosphate might require some conformational changes. Lys44, Lys46 and the conserved Trp74 are clustered at one side of the bowl where they may interact with the glycerol-3-phosphate (Figure 5). The bowl is open to solvent at its top, but the surface representation suggests that Arg113 and other residues may have to move in order to admit CTP to its binding site. The ^{113}RTEGISTT loop and helix E, the 3_{10} segment, helix B, and the loop that carries Trp74 all contribute to the rim of the substrate-binding bowl. These regions appear to be somewhat mobile as evidenced by higher than average B factors, but the ways in which they might change conformation in response to binding of substrates remain to be explored.

Figure 5

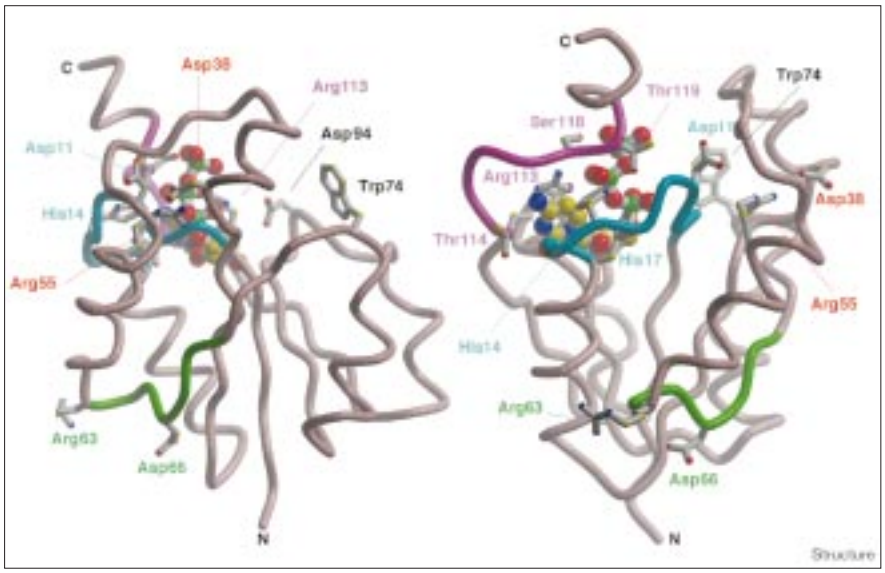


Stereoview of the substrate-binding bowl looking from the 'top' of the molecule (see Figure 2d), with dot surfaces at radii of 1.4 Å surrounding all atoms. The ^{113}RTEGISTT sequence (magenta) can be seen at the top; the ^{8}GFTDLLHWGH sequence (cyan) that connects strand 1 to the start of helix A is deeply buried at the upper left. One side of the substrate bowl is formed from these conserved sequences, while the other side (lower right) is formed from the 3_{10} segment, helix B and the loop carrying Trp74. The C termini of β-sheet strands 1 and 4 form the base of the bowl.

Mapping conserved residues onto the structure

Residues that are conserved in the GCT cytidylyltransferase family are identified in Figure 1. A number of these residues have been mutated, primarily to alanine, and the activities of the mutant enzymes measured by steady-state kinetic methods [8]. In Figure 6 the key conserved residues and the sites of most of the mutations that affect activity [8] are mapped onto the structure. Many of these sites are in the three fingerprint sequences that have already been discussed, and others, including Asp94 and Trp74, are part of the substrate-binding bowl. The conserved residues Asp11, Asp38 and Arg55 (Figure 6) do not contact CTP or contribute to the surface of the substrate-binding bowl, but mutations of these residues to alanine nevertheless decrease k_{cat}/K_m. Arg55 is the central residue of a 'hydrophilic core' that includes

Figure 6



Conserved residues mapped onto the structure. The images are related by a 90° rotation about the vertical axis. Many of the conserved residues found to be important for activity by site-directed mutagenesis are part of the substrate-binding bowl (His14, His17, Trp74, Asp94 and residues of the ^{113}RTEGISTT loop). Residues from the ^{63}RYVDEVI region are distant from the substrate-binding bowl, with Arg63 at the dimer interface.

Asp11 and Asp38 and links helix B, helix C, and the HWGH motif at the start of helix A.

Comparisons of GCT with other nucleotidyl transferases

The structure of a nucleotidyltransferase that is closely related to GCT, phosphopantetheine adenylyltransferase (PPAT) from *Escherichia coli*, was reported [33] after the submission of this paper. Like GCT, PPAT resembles the aminoacyl-tRNA synthetases and utilizes an HXGH motif, TNGH, to bind the reaction product, dephosphoCoA. Sequence alignments of GCT with PPAT reveal differences in the length and composition of the loop that precedes helix E in GCT (helix α6 in PPAT); this loop is formed by part of the RTEGISTT signature sequence in GCT. The assignment of both bacterial nucleotidyltransferases to the aminoacyl-tRNA synthetase family of enzymes is corroborated by the similarities of their folds and modes of ligand binding.

Structures have been determined for two other proteins that catalyze specialized nucleotidyl transfer reactions. The structure of CMP:2-keto-3-deoxy-manno-octonic acid synthetase (CMP-Kdo synthetase), which like GCT is specific for CTP, has been solved by Schulz and colleagues [34]. The topology of the parallel β sheet in this enzyme (3-2-1-4-6-5) is related, but not identical, to that of GCT; there are no obvious sequence motifs that are shared with GCT. The reported structure is the apo form of CMP-Kdo synthetase; comparisons of the CTP-binding sites will be necessary to assess fully the relationship between these two enzymes. Kanamycin nucleotidyltransferase catalyzes a similar reaction to GCT, but binds its substrate, ATP, in a conformation unlike that found for CTP in GCT. The enzyme adopts a fold that includes a helix bundle and a mixed β sheet [24] and is thus very different from GCT.

Assignment of GCT to the class I aminoacyl-tRNA synthetase family of folds

Searches of the database of three-dimensional structures for folds related to GCT, using DALI [35], detect members of the DNA polymerase and aminoacyl-tRNA synthetase families that resemble GCT. Similarities to class I aminoacyl-tRNA synthetases were anticipated and are examined in detail below. Among the known DNA polymerases, *Taq* polymerase displays similarity to GCT, but the matching α/β domain of *Taq* is part of the 5′→3′exonuclease site rather than the polymerase site [36]. Several other folds that do not possess nucleotidyltransferase activity also resemble GCT (see Materials and methods section). These include ribonuclease H [37] along with electron-transfer flavoprotein (ETF) [38] and other nucleotide-binding proteins such as photolyase [39], alcohol dehydrogenase [40], GMP synthase [41], and biotin carboxylase [42].

The tree of domain folds constructed by DALI groups together a closely related subset of these structures that includes ETF, the aminoacyl-tRNA synthetases, and photolyase. Comparisons of the modes of nucleotide binding in this group lead us to assign GCT to the family of aminoacyl-tRNA synthetases, as proposed by Bork *et al.* [18]. Although the Z-score for matching GCT to the A-domain of ETF is somewhat larger than for glutaminyl-tRNA synthetase, the FAD prosthetic group of ETF is bound not by the A-domain but rather by the B-domain, where the FAD phosphates interact in a fashion characteristic of α/β flavoprotein sites [28–30]. The interactions of nucleotide phosphates in GCT differ from those found in ETF but, as shown in Figure 7, clearly resemble those in the class I aminoacyl-tRNA synthetases.

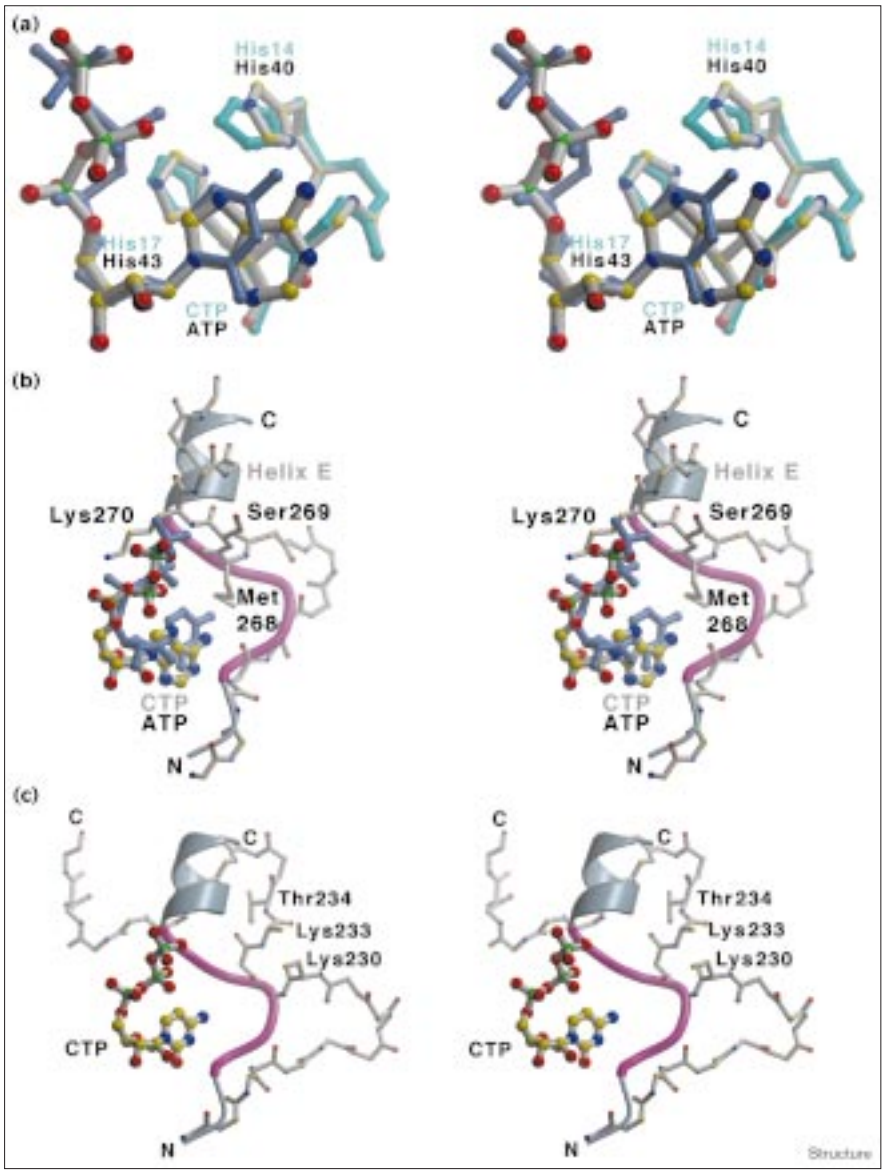
Functional and structural comparisons of GCT with class I aminoacyl-tRNA synthetases

GCT and the aminoacyl-tRNA synthetases catalyze analogous reactions at the α-phosphates of nucleoside triphosphates. The first step in the synthetase reaction is the nucleophilic attack of an amino acid carboxylate oxygen on the α-phosphate of ATP to form aminoacyl-AMP and pyrophosphate. The occurrence of the HXGH sequence in cytidylyltransferases, and the dramatic loss of activity associated with mutation of these histidines, led to the proposal that GCT, like the aminoacyl-tRNA synthetases, catalyzes nucleotidyl transfer by stabilizing the pentavalent transition state [8].

Now it is possible to compare the GCT and class I aminoacyl-tRNA synthetase structures and the way in which they bind substrates. The nucleoside triphosphate binding domains of tyrosyl- [43], glutaminyl- [44], and glutamyl-tRNA synthetases [45] match most of the fold of GCT, with which they share a common topology. Despite the low level of sequence identity (13%), 120 Cα atoms of the ATP-binding region of glutaminyl-tRNA synthetase superimpose on GCT with a root mean square deviation (rmsd) of 2.8 Å. Helices A, C, D and E of GCT correspond to comparable helices in glutaminyl-tRNA synthetase, and the 3_{10} segment of GCT occupies the same position as a more regular α helix in the synthetase.

The structural analogies extend beyond the folding and topology to local features implicated in binding and catalysis, notably the HXGH motif, which has been shown by mutation to be essential for catalysis in both the tRNA synthetases [12,46,47] and GCT [8]. The structural relationships are best seen by detailed comparison of the GCT–CTP complex with the glutaminyl-tRNA synthetase–ATP complex (Figure 7a). Significantly, the horseshoe-shaped conformation of the bound nucleotides and several of the interactions that bind the substrate phosphates to the proteins are common to both structures. There are also a few differences; for example, His40, the equivalent to His14 of GCT, is farther from the β-phosphate in glutaminyl-tRNA synthetase and interacts via a water bridge.

Figure 7



Comparisons of GCT with the class I aminoacyl-tRNA synthetases. (a) Stereoview superposition of the HXGH regions and nucleoside triphosphates (CTP and ATP, respectively) of GCT and glutaminyl-tRNA synthetase. GCT is drawn in cyan and CTP in blue. To show the similarities in the conformations and interactions of the nucleotides, the superposition was based on the ribose atoms and the Cα atoms of the HXGH sequence. Black labels refer to residues from the synthetases. (b) Stereoview comparing the bound CTP and ^{113}RTEGISTT region of GCT (blue, grey and magenta) with the ^{267}VMSKR region of glutaminyl-tRNA synthetase in complex with ATP (coordinates 1gtr). Three residues are inserted in the synthetase between the positions equivalent to Arg113 and Ile117 of GCT. As a result, this part of the chain in the synthetase loops out relative to GCT. Residues from the synthetase are drawn in thin ball-and-stick format. (c) Stereoview contrasting the bound CTP and ^{113}RTEGISTT region of GCT (grey and magenta) with the ^{230}KFGKT region of substrate-free tyrosyl-tRNA synthetase (coordinates 2ts1). The region of the synthetase corresponding to helix E of GCT is not helical, and the chain preceding the signature sequence KFGKT does not resemble the corresponding loop from the glutaminyl-tRNA synthetase (b).

Both GCT and aminoacyl-tRNA synthetases utilize a second fingerprint sequence in substrate binding and catalysis: RTXGISTT in the GCT family and the KMSKS signature sequence in aminoacyl-tRNA synthetases (KFGKT in tyrosyl-tRNA synthetase and VMSKR in glutaminyl-tRNA synthetase). These regions are less similar to one another than the HXGH motifs (Figure 7b). Without the structures it was difficult to align the RTEGISTT motif of GCT with the KMSKS motif of the synthetases, and the ^{113}RTEGISTT region was only tentatively positioned in the structure predicted by Bork *et al.* [18]. Superpositions of the structures of GCT and glutaminyl-tRNA synthetase now show that ^{117}IST and ^{231}MSK are the structurally equivalent residues. These sequences are located at the start of a helix (the C-terminal helix E in GCT) that binds the β- and γ-phosphates of the nucleotide substrates. Comparisons of the known synthetase structures suggest that interactions with substrates induce conformational changes in the KMSKS region (Figure 7c). In substrate-free tyrosyl-tRNA synthetase, the ^{230}KFGKT sequence does not form a helix whereas this region is helical in glutaminyl-tRNA synthetase complexed with ATP and tRNA. Perona *et al.* [13] propose that interaction with tRNA rather than ATP is critical for formation of the helix started by the SKS sequence. Fersht and coworkers conclude from their analyses of mutants that movements of the ^{230}KFGKT region must occur during the formation of tyrosyl adenylate [46,48,49]

and that the residues Lys230, Lys233 and Thr234 from this sequence act together to stabilize the transition state [50].

The structure of GCT strongly supports the notion that GCT, like the class I aminoacyl-tRNA synthetases, relies on positioning of substrates and transition-state stabilization for catalysis. Similarities in substrate binding and the importance of the HXGH motif for activity are consistent with similar roles for this motif in GCT and the aminoacyl-tRNA synthetases. A key question concerns the role of the fingerprint region RTEGISTT in binding and catalysis by GCT. This part of the GCT structure is shorter than the related regions in tyrosyl- or glutaminyl-tRNA synthetase but may, nevertheless, move to stabilize the transition state. Alternatively, this region may adopt different conformations in the presence of ligands other than CTP. Mutations of Ser118 and Thr119 in the ^{116}GISTT region of GCT [8] have much smaller effects than mutations at the corresponding ^{230}KFGKT residues in tyrosyl-tRNA synthetase, and it remains to be seen if Lys44 or Lys46 can be recruited to stabilize the transition state in GCT.

The dimer interface in GCT is a feature not shared with the aminoacyl-tRNA synthetases, and the structure of this interface raises the possibility of functional interactions between active centers of the dimer. The interface is extensive, with entrapped water molecules and residue interactions that include stacking, apolar contacts, and hydrogen bonding. It will be intriguing to see whether these interactions maintain a rigid interface or are altered when different ligands bind. Structure analyses of the substrate-free enzyme and of complexes of GCT with glycerol-3-phosphate, CDP-glycerol, or analogs will be useful in assessing the nature and importance of both intrachain and interchain conformational changes in the reaction cycle of GCT.

Biological implications

Cytidylyltransferases are critical enzymes involved in the biosynthetic pathways of lipids and complex carbohydrates. These enzymes catalyze a major step of energy input into biosynthesis by forming the activated intermediates, CDP-alcohols and CMP-sugars. Several cytidylyltransferases belong to a single family of structures, as defined by sequence similarities and signature sequences that occur in their catalytic domains. Among these enzymes is CTP:phosphocholine cytidylyltransferase, which regulates the synthesis of the major eukaryotic phospholipid, phosphatidylcholine. A representative member of this enzyme family is CTP:glycerol-3-phosphate cytidylyltransferase (GCT) from *Bacillus subtilis*. Because GCT consists of only the catalytic domain, it is an excellent choice for initial structure-function studies of the family.

We have solved the crystal structure of dimeric GCT to 2.0 Å resolution. Each monomer of the homodimer consists of a core α/β fold; helices and loops extending from the central β sheet form an active-site bowl in which the substrate CTP is bound. The structure reveals that highly conserved critical residues, including residues from the signature sequences HXGH and RTEGISTT, are involved in binding CTP to the active site. Conserved residues and signature sequences are also involved in monomer–monomer interactions, implicating the dimer interface as an important feature of the GCT family.

Significant similarities exist between the structure of GCT and the class I aminoacyl-tRNA synthetases, and lead us to assign the GCT fold to the superfamily of catalytic domains from class I synthetases. The protein–NTP interactions, particularly those involving the HXGH motif, are analogous in the synthetases and GCT. These similarities imply substantial mechanistic similarity. The aminoacyl-tRNA synthetases catalyze the formation of an aminoacyl-adenylate intermediate through stabilization of a pentacoordinate transition state. By confirming the similarity between the nucleotide-binding sites of these two classes of enzymes, the structure supports the proposal that the catalytic mechanisms are also analogous.

The structure of GCT is the first structure of any member of this cytidylyltransferase family and prompts further studies of the reaction mechanism. It will be important to determine whether interactions across the dimer interface have a role in the reaction catalyzed by GCT. Furthermore, this structure represents a starting point for studies on the regulation of activity in the more complex eukaryotic cytidylyltransferases.

Materials and methods

Crystallization

Recombinant GCT protein was purified from the overexpressing HMS174(DE3)pLysS strain of *E. coli* cells by blue-sepharose chromatography [15]. Crystals of GCT were grown by the hanging-drop method at 4°C. Monoclinic crystals grew within one to two weeks after mixing protein solution (10 mg/ml in 10 mM Tris pH 8.5, 1 mM EDTA, 1 mM dithiothreitol [DTT]) with a precipitant solution containing 30% polyethylene glycol (PEG) 3350 (Baker), 200 mM lithium sulfate, 100 mM Tris, pH 8.5. Crystal size was optimal with initial 5:3 ratios of protein to precipitant solution. Lithium sulfate was essential for crystallization and could not be replaced by the sodium or potassium salts. Rhombohedral crystals grew after two to three weeks of equilibration using 20% PEG 8000 (EM Sciences), 300 mM lithium sulfate, 50 mM Tris, pH 7.9 as the precipitant and well solution. 1 mM DTT was included in the well solutions in both crystallization protocols.

X-ray measurements

Native data sets were initially collected from monoclinic crystals at 4°C and from rhombohedral crystals at 120K, using a San Diego multiwire area detector system. For data collection at low temperatures, rhombohedral crystals were stabilized in well solution that included 20% (v/v) PEG 200 (Fluka). Derivative crystals were prepared by soaking native crystals overnight in 10 mM mercuric iodide dissolved in well solution. Intensities were integrated and corrected using the SDMS software [51].

Table 1

Data collection statistics.

Data set	Resolution (Å)	Wavelength (Å)	Total reflections	Unique reflections	Completeness* (%)	R_{sym}* (%)
Native ($P2_1$)	2.0	1.5418	45,342	16,917	88.6 (83.8)	6.1 (12.5)
Native (R3)	3.2	1.5418	11,096	5,189	98.1 (96.7)	6.0 (12.3)
Hg-1 (R3)	2.9	1.0088	106,456	12,695†	97.4 (99.5)	7.3 (11.1)
Hg-2 (R3)	2.9	1.0067	107,278	12,745†	97.2 (99.5)	7.1 (10.9)
Hg-3 (R3)	2.9	0.97638	103,511	12,551†	96.8 (99.7)	7.2 (11.6)
Hg (R3)	3.15	1.5418	9,115	4,359	79.0 (68.9)	6.9 (12.3)

*Values in parentheses are for the highest resolution shell. †Bijvoet pairs separated.

MAD measurements on the mercuric iodide derivative of the R3 crystal form were carried out at beamline F2 at CHESS. The crystal was cooled to 100K in the cryoprotectant solution described above. Diffraction data were collected at the inflection point (12.292 keV), the peak (12.317 keV) and a remote wavelength (12.700 keV). The MAD data sets were processed using DENZO and SCALEPACK [52].

Structure determination

Initial phases for the rhombohedral GCT structure were determined using the multiwavelength measurements from the mercury-substituted crystals. Difference Patterson maps, calculated with data from native GCT and the mercuric iodide derivative, had revealed a single dominant mercury-binding site in the asymmetric unit of the R3 crystal form, and the parameters for this site were refined in HEAVYREF [53]. MAD phasing was then carried out with MLPHARE [54], using multiwire detector data from an HgI_2 derivative crystal along with the data sets collected at CHESS. For phasing, the inflection point wavelength was used as the native data set and the three other wavelengths were treated as separate derivatives. Maps calculated to 3.5 Å resolution with an estimated < m > of 0.57 revealed most of the α/β core of the structure. Subsequent phase modification, including solvent flattening and histogram matching, was accomplished using DM [54] and with extension of phasing to 3.0 Å led to maps that permitted identification of the noncrystallographic symmetry (NCS) related monomers. NCS-averaging was performed with DM [54].

Twinning correction

The analysis of twinning in the monoclinic crystal form will be described in detail elsewhere (CHW and MLL, unpublished results). Twinning in these crystals arises from a 180° rotation about the a axis and is merohedral (i.e. superimposes the twin-related reflections). The data used for refinement of the monoclinic form (Table 1) were detwinned after determining the twin fraction for the native crystal to be 0.13 according to the method of Britton [55].

Refinement

The structure of the rhombohedral form of GCT was refined using the program X-PLOR [56]. Because of the low data/parameter ratios at the available resolution of 2.8 Å, strict NCS constraints were imposed and B-factor shifts were determined for mainchain and sidechain groups. After several rounds of refinement, including simulated annealing, the rhombohedral GCT model included 252 of 258 residues and two CTP molecules, with an R_{cryst} of 0.323 and R_{free} of 0.383 (5% of the reflections) for all data from 8.0 Å to 2.8 Å.

The molecular model of GCT described in detail in this paper has been derived by subsequent refinement with data from the monoclinic $P2_1$ crystal form (Table 1). The initial model was obtained by molecular replacement from the structure of the rhombohedral form using X-PLOR [56]. Rotation searches produced a peak at 6σ above background, and the subsequent translation function produced a peak at 23σ. The R factor for this solution was 0.467 for all data from 6.0 Å to 3.0 Å.

Refinement started with the oriented search model that included residues 1–126 of both monomers. The monoclinic data, from one native crystal, had been detwinned (see above). Four rounds of computations that included positional refinement, simulated annealing, and adjustment of thermal factors were alternated with rebuilding based on omit maps. In the omit calculations, which used simulated annealing, the entire model was examined in stepwise fashion, omitting approximately 10% of the residues in each individual calculation. After these and further refinements in X-PLOR [56], the final model of GCT based on data from the monoclinic form includes 252 of the 258 residues in the dimer, two CTP molecules, and 114 solvent molecules. R_{cryst} and R_{free} are 0.196 and 0.255, respectively, for all observed data from 8.0 Å to 2.0 Å (Table 2). Calculation of the local symmetry, using the program O [57], shows that the two polypeptide chains in the asymmetric unit are superimposed by a rotation of 179.5° and a translation of 0.09 Å. To re-examine the effects of twinning [58], additional refinement has been performed with SHELXL-97 [59] using the native data without twinning corrections. The twinning fraction (α) can be refined using SHELXL in such a way that only one additional parameter is introduced, and was determined to be 0.209, somewhat higher than derived from the Britton plots.

During adjustment of the model, solvent molecules were included if a 3σ peak was present in ($|F_o|-|F_c|$) difference maps, and retained if the B factor at full occupancy remained below 55 $Å^2$. The three C-terminal residues have been omitted from the model. Sidechains of several surface residues cannot be modeled completely, and the electron density suggests multiple conformations at Ser36, Ile62 and Ile81. PROCHECK [60] was used to assess the quality of the final model. The model shows good stereochemistry for a 2.0 Å structure; 90.1% of the residues fall in the most favored regions of the Ramachandran map and no residues are found in disallowed regions.

Table 2

Data refinement statistics.

Resolution (Å)	8.0–2.0
R_{work}	0.196*
R_{free}	0.255
No. residues (total)	252 (258)
No. nonhydrogen atoms	2294 (per dimer)
No. water molecules	111
Refined <B> ($Å^2$)	25.2
Wilson <B> ($Å^2$)	27.2
Rmsd bonds (Å)	0.010
Rmsd angles (°)	1.472
Rmsd impropers (°)	1.358
Ramachandran plot	
most favored (%)	91.8
additional (%)	8.2
disallowed (%)	0.0

*No σ cut-off.

Database searches and classification of the GCT fold

To identify structures related to GCT, comparisons of GCT with a database of 1396 proteins of known three-dimensional structure were carried out using DALI [35]. Fourteen α/β proteins with Z-scores greater than 4.5 were detected. The α/β domain of the A-domain of ETF [38], with a Z-score of 8.3, possesses the highest degree of structural similarity to GCT. Three class I aminoacyl-tRNA synthetases (tyrosyl- [43], glutaminyl- [44] and glutamyl-tRNA synthetases [45]) aligned with Z-scores greater than 7.0. T4 RNase H [37] and the N-terminal exonuclease domain of *Taq* DNA polymerase [36] also resemble GCT. Other related folds are found in photolyase [39], biotin carboxylase [42], GMP synthetase [41], folylpolyglutamate synthase [61], enoyl-acyl carrier protein [60], alcohol dehydrogenase [40], and methionyl-tRNA f-Met formyltransferase [62]. The hierarchy of folds obtained from the DALI web site (www.embl-ebi.ac.uk/dali) was also used to assess relationships among the folds that resemble GCT.

Generation of figures

Figures 2, 3, 6 and 7 were generated using the program MOLSCRIPT [63] and the rendering program VORT (University of Melbourne). Figure 5 was drawn using RIBBONS [64] and Figure 4 was made with LIGPLOT [60].

Accession numbers

The coordinates of GCT have been deposited in the Protein Data Bank and assigned accession number 1coz.

Acknowledgements

We would like to thank Erik Zuiderweg and Eric Schurter for helpful discussions. Synchrotron data were collected at CHESS beamline F2, thanks to Dan Thiel and Marian Szebenyi. This research has been supported by NIH grants CA 64159 (CK), GM 16429 (MLL) and a training grant (GM 07315) in cellular and molecular biology (CHW).

References

1. Cornell, R. (1989). Chemical cross-linking reveals a dimeric structure for CTP: phosphocholine cytidylyltransferase. *J. Biol. Chem.* **264**, 9077-9082.
2. Kent, C. (1997). CTP: phosphocholine cytidylyltransferase. *Biochim. Biophys. Acta* **1348**, 79-90.
3. Cornell, R. & Vance, D.E. (1987). Translocation of CTP: phosphocholine cytidylyltransferase from cytosol to membranes in HeLa cells: stimulation by fatty acid, fatty alcohol, mono- and diacylglycerol. *Biochim. Biophys. Acta* **919**, 26-36.
4. Weinhold, P.A., Rounsifer, M.E., Charles, L. & Feldman, D.A. (1989). Characterization of cytosolic forms of CTP: choline-phosphate cytidylyltransferase in lung, isolated alveolar type II cells, A549 cell and Hep G2 cells. *Biochim. Biophys. Acta* **1006**, 299-310.
5. Skinner, H.B., *et al.*, & Bankaitis, V.A. (1995). The *Saccharomyces cerevisiae* phosphatidylinositol-transfer protein effects a ligand-dependent inhibition of choline-phosphate cytidylyltransferase activity. *Proc. Natl Acad. Sci. USA* **92**, 112-116.
6. Baburina, I. & Jackowski, S. (1998). Apoptosis triggered by 1-O-octadecyl-2-O-methyl-rac-glycero-3-phosphocholine is prevented by increased expression of CTP:phosphocholine cytidylyltransferase. *J. Biol. Chem.* **273**, 2169-2173.
7. MacDonald, J.I. & Kent, C. (1994). Identification of phosphorylation sites in rat liver CTP:phosphocholine cytidylyltransferase. *J. Biol. Chem.* **269**, 10529-10537.
8. Park, Y.-S., *et al.*, & Kent, C. (1997). Identification of functional conserved residues of CTP:glycerol-3-phosphate cytidylyltransferase: role of histidines in the conserved HXGH in catalysis. *J. Biol. Chem.* **272**, 15161-15166.
9. Min-Seok, R., Kawamata, Y., Nakamura, H., Ohta, A. & Takagi, M. (1996). Isolation and characterization of ECT1 gene encoding CTP: phosphoethanolamine cytidylyltransferase of *Saccharomyces cerevisiae*. *J. Biochem.* **120**, 1040-1047.
10. Nakashima, A., Hosaka, K. & Nikawa, J. (1997). Cloning of a human cDNA for CTP-phosphoethanolamine cytidylyltransferase by complementation *in vivo* of a yeast mutant. *J. Biol. Chem.* **272**, 9567-9572.
11. Delarue, M. & Moras, D. (1993). The aminoacyl-tRNA synthetase family: modules at work. *BioEssays* **15**, 675-687.
12. Leatherbarrow, R.J., Fersht, A.R. & Winter, G. (1985). Transition-state stabilization in the mechanism of tyrosyl-tRNA synthetase revealed by protein engineering. *Proc. Natl Acad. Sci. USA* **82**, 7840-7844.
13. Perona, J.J., Rould, M.A. & Steitz, T.A. (1993). Structural basis for transfer RNA aminoacylation by *Escherichia coli* glutaminyl-tRNA synthetase. *Biochemistry* **32**, 8758-8771.
14. Rossmann, M.G., Liljas, A., Bränden, C.-I. & Banaszak, L.K. (1975). Evolutionary and structure relationships among dehydrogenases. In *The Enzymes Vol. 11.* (Boyer, P.D., ed.), pp. 61-102, Academic Press, New York, USA.
15. Park, Y.-S., Sweitzer, T.D., Dixon, J.E. & Kent, C. (1993). Expression, purification, and characterization of CTP:glycerol-3-phosphate cytidylyltransferase from *Bacillus subtilis*. *J. Biol. Chem.* **268**, 16648-16654.
16. Jones, S. & Thornton, J.M. (1995). Protein–protein interactions: a review of protein dimer structures. *Prog. Biophys. Mol. Biol.* **63**, 31-65.
17. Sweitzer, T.D. & Kent, C. (1994). Expression of wild type and mutant rat liver CTP:phosphocholine cytidylyltransferase in a cytidylyltransferase deficient Chinese hamster ovary cell line. *Arch. Biochem. Biophys.* **311**, 107-116.
18. Bork, P., Holm, L., Koonin, E.V. & Sander, C. (1995). The cytidylyltransferase superfamily: identification of the nucleotide-binding site and fold prediction. *Proteins* **22**, 259-266.
19. Brändén, C.I. (1980). Relation between structure and function of α/β proteins. *Quart. Rev. Biophys.* **13**, 317-338.
20. Berry, M.B., *et al.*, & Phillips, G.N., Jr. (1994). The closed conformation of a highly flexible protein: the structure of *E. coli* adenylate kinase with bound AMP and AMPPNP. *Proteins* **19**, 183-198.
21. Abele, U. & Schulz, G.E. (1995). High-resolution structures of adenylate kinase from yeast ligated with inhibitor Ap5A, showing the pathway of phosphoryl transfer. *Protein Sci.* **4**, 1262-1271.
22. Abrahams, J.P., Leslie, A.G.W., Lutter, R. & Walker, J.E. (1994). Structure at 2.8 Å resolution of F_1-ATPase from bovine heart mitochondria. *Nature* **370**, 621-628.
23. Doublie, S., Tabor, S., Long, A.M., Richardson, C.C. & Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* **391**, 251-258.
24. Pedersen, L.C., Benning, M.M. & Holden, H.M. (1995). Structural investigation of the antibiotic and ATP-binding sites in kanamycin nucleotidyltransferase. *Biochemistry* **34**, 13305-13311.
25. Walker, J.E., Saraste, M., Runswick, M.J. & Gay, N.J. (1982). Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**, 945-951.
26. Saraste, M., Sibbald, P.R. & Wittinghofer, A. (1990). The P-loop – a common motif in ATP- and GTP-binding proteins. *Trends Biochem. Sci.* **15**, 430-434.
27. Traut, T.W. (1994). The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites. *Eur. J. Biochem.* **222**, 9-19.
28. Wierenga, R.K., Drenth, J. & Schulz, G. (1983). Comparison of the three-dimensional protein and nucleotide structure of the FAD-binding domain of *p*-hydroxybenzoate hydroxylase with the FAD- as well as NADPH-binding domains of glutathione reductase. *J. Mol. Biol.* **167**, 725-739.
29. Eggink, G., Engel, H., Vriend, G., Terpstra, P. & Witholt, B. (1990). Rubredoxin reductase of *Pseudomonas oleovorans*. Structural relationship to other flavoprotein oxidoreductases based on one NAD and two FAD fingerprints. *J. Mol. Biol.* **212**, 135-142.
30. Schulz, G. (1992). Binding of nucleotides by proteins. *Curr. Opin. Struct. Biol.* **2**, 61-67.
31. Kennedy, E.P., Borkenhage, L.F. & Smith, S.W. (1959). Possible metabolic functions of deoxycytidine diphosphate choline and deoxycytidine diphosphate ethanolamine. *J. Biol. Chem.* **234**, 1998-2000.
32. Sankar, S., Ballou, D.P. & Kent, C. (1998). The K_d values for binding of all substrates to CTP: glycerol-3-phosphate cytidylyltransferase are about 1000-fold less than the K_m values. *FASEB J.* **12**, A1287.
33. Izard, T. & Geerlof, A. (1999). The crystal structure of a novel bacterial adenylyltransferase reveals half of sites reactivity. *EMBO J.* **18**, 2021-2030.
34. Jelakovic, S., Jann, K. & Schulz, G.E. (1996). The three-dimensional structure of capsule-specific CMP: 2-keto-3-deoxy-manno-octonic acid synthetase from *E. coli*. *FEBS Lett.* **391**, 157-161.
35. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138.

36. Eom, J.H., Wang, J. & Steitz, T.A. (1996). Structure of *Taq* polymerase with DNA at the polymerase active site. *Nature* 382, 278-281.
37. Mueser, T.C., Nossal, N.G. & Hyde, C.C. (1996). Structure of bacteriophage T4 RNase H, a 5′ to 3′ RNA-DNA and DNA-DNA exonuclease with sequence similarity to the RAD2 family of eukaryotic proteins. *Cell* 85, 1101-1112.
38. Roberts, D.L., Frerman, F.E. & Kim, J.P. (1996). Three-dimensional structure of human electron transfer flavoprotein to 2.1 Å resolution. *Proc. Natl Acad. Sci. USA* 93, 14355-14360.
39. Tamada, T., *et al*., & Miki, K. (1997). Crystal structure of DNA photolyase from *Anacystis nidulans*. *Nat. Struct. Biol.* 4, 887-891.
40. Ramaswamy, S., Eklund, H. & Plapp, B.V. (1994). Structures of horse liver alcohol dehydrogenase complexed with NAD^+ and substituted benzyl alcohols. *Biochemistry* 33, 5230-5237.
41. Tesmer, J.J.G., Klem, T.J., Deras, M.L., Davisson, V.J. & Smith, J.L. (1996). The crystal structure of GMP synthetase reveals a novel catalytic triad and is a structural paradigm for two enzyme families. *Nat. Struct. Biol.* 3, 74-86.
42. Waldrop, G.L., Rayment, I. & Holden, H.M. (1994). Three-dimensional structure of the biotin carboxylase subunit of acetyl-CoA carboxylase. *Biochemistry* 33, 10249-10256.
43. Brick, P., Bhat, T.N. & Blow, D.M. (1989). Structure of tyrosyl-tRNA synthetase refined at 2.3 Å resolution. Interaction of the enzyme with the tyrosyl adenylate intermediate. *J. Mol. Biol.* 208, 83-98.
44. Rould, M.A., Perona, J.J., Soll, D. & Steitz, T.A. (1989). Structure of *E. coli* glutaminyl-tRNA synthetase complexed with tRNA(Gln) and ATP at 2.8 Å resolution. *Science* 246, 1135-1142.
45. Nureki, O., *et al*., & Morikawa, K. (1995). Architectures of class-defining and specific domains of glutamyl-tRNA synthetase. *Science* 267, 1958-1965.
46. Fersht, A.R., Knill-Jones, J.W., Bedouelle, H. & Winter, G. (1988). Reconstruction by site-directed mutagenesis of the transition state for the activation of tyrosine by the tyrosyl-tRNA synthetase: a mobile loop envelopes the transition state in an induced-fit mechanism. *Biochemistry* 27, 1581-1587.
47. Lowe, D.M., Fersht, A.R., Wilkinson, A.J., Carter, P. & Winter, G. (1985). Probing histidine-substrate interactions in tyrosyl-tRNA synthetase using asparagine and glutamine replacements. *Biochemistry* 24, 5106-5109.
48. Fersht, A.R. (1988). Relationships between apparent binding energies measured in site-directed mutagenesis experiments and energetics of binding and catalysis. *Biochemistry* 27, 1577-1580.
49. Avis, J.M. & Fersht, A.R. (1993). Use of binding energy in catalysis: optimization of rate in a multistep reaction. *Biochemistry* 32, 5321-5326.
50. First, E.A. & Fersht, A.R. (1995). Analysis of the role of the KMSKS loop in the catalytic mechanism of the tyrosyl-tRNA synthetase using multimutant cycles. *Biochemistry* 34, 5030-5043.
51. Howard, A.J., Nielsen, C. & Xuong, N.H. (1985). Software for a diffractometer with multiwire area detector. *Methods Enzymol.* 114, 452-472.
52. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276, 307-326.
53. Terwilliger, T.C. & Eisenberg, D. (1983). Unbiased three-dimensional refinement of heavy-atom parameters by correlation of origin-removed Patterson functions. *Acta Crystallogr. A* 39, 813-817.
54. Collaborative Computational Project No. 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* 50, 760-763.
55. Britton, D. (1972). Estimation of twinning parameter for twins with exactly superimposed reciprocal lattices. *Acta Crystallogr. A* 28, 296.
56. Brünger, A.T. (1992). *X-PLOR: a System for Crystallography and NMR*. Yale University Press, New Haven CT.
57. Jones, T.A. & Kjeldgaard, M. (1993). *O – The Manual*. Uppsala University, Uppsala, Sweden.
58. Luecke, H., Richter, H.-T. & Lanyi, J.K. (1998). Proton transfer pathways in bacteriorhodopsin at 2.3 Å resolution. *Science* 280, 1934-1937.
59. Sheldrick, G.M. & Schneider, T.R. (1997). SHELXL: high-resolution refinement. *Methods Enzymol.* 277, 319-343.
60. Wallace, A.C., Laskowski, R.A. & Thornton, J.M. (1995). LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng.* 8, 127-134.
61. Sun, X., Bognar, A.L., Baker, E.N. & Smith, C.A. (1998). Structural homologies with ATP- and folate-binding enzymes in the crystal structure of folylpolyglutamate synthetase. *Proc. Natl Acad. Sci. USA* 95, 6647-6652.
62. Schmitt, E., Panvert, M., Blanquest, S. & Mechulam, Y. (1998). Crystal structure of methionyl-tRNAfMet transformylase complexed with the initiator formyl-methionyl-tRNAfMet. *EMBO J.* 17, 6819-6826.
63. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of proteins. *J. Appl. Crystallogr.* 24, 946-950.
64. Carson, M. (1997). Ribbons. *Methods Enzymol.* 277, 493-504.

Because *Structure with Folding & Design* operates a 'Continuous Publication System' for Research Papers, this paper has been published on the internet before being printed (accessed from http://biomednet.com/cbiology/str). For further information, see the explanation on the contents page.